

Strength of Connections in a Random Graph: Definition, Characterization, and Estimation

Subhadeep Mukhopadhyay

Department of Statistics, Temple University

Philadelphia, Pennsylvania, 19122, U.S.A.

deep@temple.edu

ABSTRACT

How can the ‘*affinity*’ or ‘*strength*’ of ties of a random graph be characterized and compactly represented? How can concepts like *Fourier and inverse-Fourier* like transform be developed for graph data? To do so, we introduce a new graph-theoretic function called ‘Graph Correlation Density Field’ (or in short **GraField**), which differs from the traditional edge probability density-based approaches, to completely characterize tie-strength between graph nodes. Our approach further allows frequency domain analysis, applicable for both directed and undirected random graphs.

Keywords: Piecewise constant orthonormal LP Graph basis; LP Graph transform; Graph correlation density field; Graphon.

1 Introduction

Lovász and Szegedy (2006) introduced the concept of Graphon, as a functional modeling tool to compactly describe the edge probabilities of a random graph. Characterization and estimation based on graphon is now a fast-growing branch of (non-parametric) statistics and graph theory; see for instance Airoldi et al. (2013), Olhede and Wolfe (2014), Chatterjee (2014), Frieze and Kannan (1999), and Lovász (2012).

However, the fundamental quantity of interest in graphs or networks is *not* the ‘edge probabilities,’ rather it is the connections (or interaction) between pairs of vertices.

In the literature of random graph modeling there *has not been any attempt* to define such ‘strength’ of connections (or interaction) between pairs of vertices. In this paper, our aim is to construct a function that characterizes this *connectedness or tie-strength* (instead of of representing the probability of a tie or edge between any two nodes) in a compact way and to study its fundamental role in random graph modeling. The central goal of this paper is to answer:

How can such implicit affinity function be constructed and estimated from explicitly observed links?

We introduce a fundamental object, called ‘Graph Correlation Density Field’ (**GraField**), as a tool to provide a compact functional description of graph affinity (how related any two vertices are), formulated precisely in Section 2.3 and 2.4. Orthogonal decomposition of this function is also derived to efficiently represent and nonparametrically estimate the underlying correlation field. We also address the problem of graph *compression* and *sparse* representation learning to model the interactions between large numbers of nodes - a major research challenge in modern graph theory.

Furthermore, we show that how the concept of **GraField** provides a systematic and comprehensive framework for frequency-domain analysis of random graphs, which is *analogous* to its time series counterparts. In particular, we are interested in the following question, which plays a fundamental role in graph signal processing:

How can concepts like of Fourier and inverse-Fourier like transform be developed for graph data?

Our nonparametric approximation scheme of **GraField** hinges upon a specially designed orthogonal transform basis. Section 2.1 introduces the theory and construction of this new kind of discrete (piecewise constant) orthonormal basis set called LP Graph basis. We define the concept of LP graph transforms (analogue of Fourier transform in time series) in Section 2.2 to allow a frequency interpretation of graph data parallel to time series analysis that captures graph connectivity patterns and

plays a vital role in compression. We also show that the LP graph transform coefficients can be interpreted as discrete orthogonal transforms of the fundamental Graph Correlation Density function.

We apply this new graph correlation field theory to solve two graph modeling problems in Section 3. The first step of modeling is to identify the null model. In time series analysis, the ‘flat’ shape of power spectral density and the plot auto-correlations provide an elegant diagnostic tool for white noise process (which is the null model in the context of time series). Section 3.1 addresses the following question:

Can we develop an analogous portmanteau graphical diagnostic test (similar in spirit to time series) for null random graph models?

We discuss two graphical diagnostic tests that resemble the sample auto-correlation function and spectral density-based approach for white noise testing in times series context. We show that for a null graph model, the shape of ‘Graph Correlation Density’ is ‘flat’ and the sample LP-graph transforms behave asymptotically normal distribution with mean 0 and variance $1/N$, remarkably similar to its time series counterparts (Brockwell and Davis, 2009).

Section 3.2 explores the connection between two fundamental functions of modern graph theory: Graph Correlation Density and the traditional edge probability-based concept, Graphon (Lovász, 2012). As a practical utility of this connection we answer the following question:

Can we develop a sparse estimation strategy that will provide ‘smooth’ nonparametric estimate of Graphon?

Our research advances the current state-of-the-art nonparametric graphon modeling (that yields histogram-based approximation) by offering a low-rank efficient representation-based smoothing algorithm. We end with final conclusions and future research directions in Section 4.

2 Theory and Estimation

2.1 Discrete Basis Construction

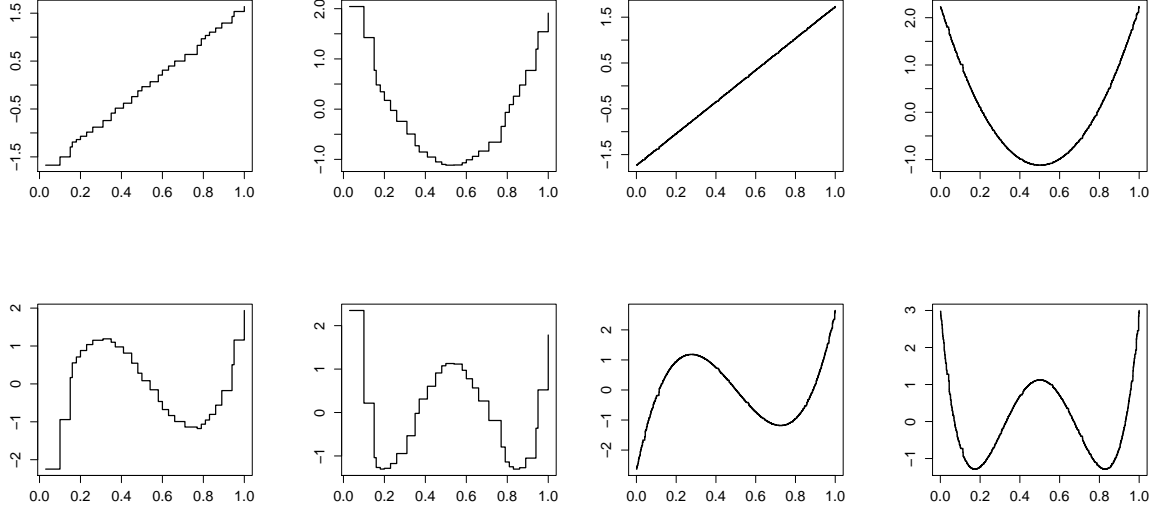


Figure 1: (a) Left: Piecewise constant shapes of first four LP orthonormal basis for the bipartite graph with $n = 30$ (divided equally into two disjoint subsets) and $p = .25$; (b) Right: first four basis functions for political blog data that resemble orthonormal shifted Legendre polynomials on $(0, 1)$.

The traditional approach constructs nonparametric orthonormal graph basis by performing singular-value decomposition (SVD) of Laplacian or Adjacency matrix (both matrices are of size $n \times n$, where n is the graph size). The question that naturally arises is as follows:

Can we bypass this computationally expensive SVD of a (possibly) large matrix to find the graph Fourier-like discrete orthonormal bases?

In what follows, we propose an algorithm that finds nonlinear adaptive discrete graph basis avoiding SVD altogether - computationally fast. Our basis is nonparametric (*not fixed* like discrete Fourier basis, wavelets, or curvelets, etc.) in the sense that its shape depends on (or adapts to) the given graph structure. Furthermore, these custom design basis functions will act as an expansion basis for our functional graph

object $\mathcal{C}(u, v; \mathcal{G})$ ‘graph correlation density’.

Here, we provide the recipe for how to construct discrete piecewise constant orthonormal basis functions, specially designed for the given random graph. Let us start with the construction of orthonormal polynomials associated with an arbitrary discrete random variable X . For a given discrete random variable X , we construct the basis functions by Gram Schmidt orthonormalization of powers of $T_1(x; X)$, given by

$$T_1(x; X) = \frac{\sqrt{12}[F^{\text{mid}}(x; X) - .5]}{\sqrt{1 - \sum_x p^3(x; X)}}, \quad (2.1)$$

where

$$F^{\text{mid}}(x; X) = F(x; X) - .5p(x; X), \quad p(x; X) = \Pr[X = x], \quad F(x; X) = \Pr[X \leq x]. \quad (2.2)$$

By construction, these basis functions satisfy the following orthonormality conditions:

$$\sum_i T_j(x_i; X)p(x_i; X) = 0, \quad \sum_i T_j^2(x_i; X)p(x_i; X) = 1, \quad \sum_i T_j(x_i; X)T_k(x_i; X)p(x_i; X) = 0. \quad (2.3)$$

We define LP orthonormal basis in the *unit interval* $[0, 1]$ (rescaling operation) by evaluating T_j at $Q(u; X)$, quantile function of the random variable X

$$S_j(u; X) = T_j(Q(u; X); X), \quad 0 < u < 1. \quad (2.4)$$

For a given directed (and possibly weighted) random graph of size n , we construct two sets of nonlinear adaptive discrete orthogonal graph bases $\{S_1(u; X), S_2(u; X), \dots, S_{n-1}(u; X)\}$ and $\{S_1(v; Y), S_2(v; Y), \dots, S_{n-1}(v; Y)\}$ based on in-degree and out-degree sequences

$$p(x; X) = \sum_{y=1}^n A(x, y)/N, \quad p(y; Y) = \sum_{x=1}^n A(x, y)/N, \quad p(x, y; \mathcal{G}) = A(x, y)/N, \quad N = \sum_{x, y=1}^n A(x, y), \quad (2.5)$$

where A is the (weighted) adjacency matrix of the graph \mathcal{G} . The Eq (2.3) guarantees that both basis sets are automatically *degree-weighted*. Note that for undirected graphs we only have one set of basis due to the symmetry.

Figure 1(a) shows the ‘piecewise constant’ (nonlinear) shapes of the first four orthonormal basis functions of the bipartite graph with $n = 30$ (divided equally into two disjoint subsets) and connection probability $p = .25$. However, the LP graph bases exhibit a ‘smooth continuous shape’ for the political blog data (Adamic and Glance, 2005), as shown in Figure 1(b). This is a directed graph with $n = 1490$ and the total of directed edges (incoming and outgoing links) 19,090. This data was collected immediately after the 2004 presidential election to understand the connection patterns (as indicated by hyperlinks) between two political web-blogs (liberal and conservative).

One unique aspect of our theory is that it enables us to pass from the ‘piecewise-constant discrete’ orthonormal basis of \mathbb{R}^n (for small graphs) to an ‘approximately smooth’ continuous orthonormal basis of $L^2(0, 1)$ (for large graphs). In the *continuum limit* (as dimensional of the graph $n \rightarrow \infty$), our bases exhibit a ‘*universality*’ property in the sense that it approaches to shifted orthonormal Legendre Polynomials on $(0, 1)$. To emphasize this universal limiting shape, we call it ‘LP’ graph basis functions.

2.2 Orthogonal Graph LP Transform

We introduce a Fourier-like transform, called graph LP transform (valid for directed/undirected graphs) that provides a novel scheme for capturing the structural information of a graph using few coefficients, thus achieving the goal of graph compression.

Compute LP Graph transforms by taking cross-covariance (which can be interpreted as correlation due to (2.3), thus we can alternatively call it LP graph correlations) between the LP-basis functions $T_j(X; \mathcal{G})$ and $T_k(Y; \mathcal{G})$ for $j = \{1, 2, \dots, n-1\}$ and $k = \{1, 2, \dots, n-1\}$

$$\text{LP}[j, k; \mathcal{G}] = \mathbb{E}[T_j(X; \mathcal{G})T_k(Y; \mathcal{G})] = \sum_{x,y \in \{1,2,\dots,n\}} p(x, y; \mathcal{G})T_j(x; \mathcal{G})T_k(y; \mathcal{G}) \equiv \langle T_j, T_k \rangle_{L^2(p)}. \quad (2.6)$$

We go from the original graph domain to the LP frequency domain since important

structural information of a graph may be compressed into fewer discrete LP transform coefficients. Strictly speaking, the topology for structured graphs tends to be concentrated in a few low-frequency (low values of j and k) LP graph transforms - leading to efficient representation. Each LP graph correlations $LP[j, k; \mathcal{G}]$ captures different connection patterns of the graph. As we will see in the next section, these LP transform coefficients can be interpreted as discrete orthogonal transforms of the graph correlation density kernel.

2.3 Graph Correlation Density Field

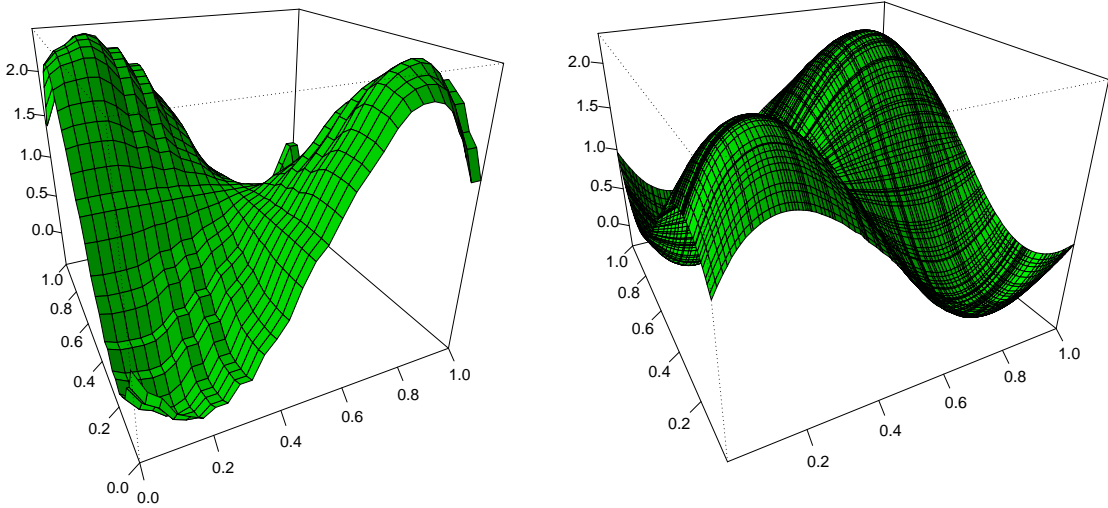


Figure 2: (color online) Nonparametric smooth graph correlation density estimate for (a) Left: bipartite graph with $n = 30$ and $p = .25$, and (b) Right: directed political blog graph data with $n = 1490$. Note the transition from bivariate step-function (small graph of size $n = 30$) to smooth continuous shape (for large graph with $n = 1490$ nodes) of the estimated GraField $\hat{\mathcal{C}}(u, v; \mathcal{G})$.

To completely characterize and compactly represent the ‘affinity’ or ‘strength’ of ties (or interaction) between every pair of vertices in the graph, we introduce a new functional statistical tool, called ‘Graph Correlation Density Field’ (**GraField**), denoted by $\mathcal{C}(u, v; \mathcal{G})$. We also discuss estimation of this (implicit) tie-strength function to capture the underlying interactions from the observed adjacency matrix

A.

Theorem 2.1. *The two equivalent dual representations of orthogonal LP graph transform is given by*

$$\text{LP}[j, k; \mathcal{G}] = \langle T_j, T_k \rangle_{L^2(p)} = \langle S_j, \int_0^1 \mathcal{C} S_k \rangle_{L^2[0,1]} \quad (j, k = 1, \dots, n), \quad (2.7)$$

where the square integrable function $\mathcal{C} : [0, 1]^2 \rightarrow \mathbb{R}_+ \cup \{0\}$ is given by

$$\mathcal{C}(u, v; \mathcal{G}) = \frac{p(Q(u; X), Q(v; Y); \mathcal{G})}{p(Q(u; X))p(Q(v; Y))}, \quad (2.8)$$

where $u = F(x; X), v = F(y; Y)$ for $x, y \in \{1, 2, \dots, n\}$.

The following fundamental representation theorem can be interpreted as *inverse LP transform* (in light of Theorem 2.1), which converts the set of LP-coefficients $\text{LP}[j, k; \mathcal{G}]$ into a density function over unit square.

Theorem 2.2. *The function \mathcal{C} defined in Eq. (2.8) for a directed graph with size n , is a proper density function that admits the following orthogonal decomposition in terms of piecewise-constant product LP-basis over unit square for $u = F(x; X), v = F(y; Y)$ and $x, y \in \{1, 2, \dots, n\}$.*

$$\mathcal{C}(u, v; \mathcal{G}) - 1 = \sum_{j=1}^{n-1} \sum_{k=1}^{n-1} \text{LP}[j, k; \mathcal{G}] S_j(u; X) S_k(v; Y), \quad (2.9)$$

This can be considered as a frequency (LP) domain functional representation of graph topology, which computes the underlying (unobserved) ‘*affinity* to connect’ (not the un-normalized ‘raw’ probability of an edge) between graph nodes. As the total area under this non-negative correlation function is one, we call it ‘density function’. LP space description of the graph structure (2.9) allows a Fourier-like interpretation. This also shows that the LP-coefficients can be used to reconstruct graph topology, captured by $\mathcal{C}(u, v)$, which can be represented as the weighted sum of custom-designed product-LP graph basis functions. Perform ‘smooth’ estimation by selecting the ‘large’ LP-components $\text{LP}[j, k; \mathcal{G}]$ in a data-driven way (discussed in the next section).

We quantify the graph structure by measuring the departure from uniformity

$$\text{LPINFOR}(\mathcal{G}) = \sum_{j=1}^{n-1} \sum_{k=1}^{n-1} \left| \text{LP}[j, k; \mathcal{G}] \right|^2 = \int_{[0,1]^2} \mathcal{C}^2 - 1 \quad (2.10)$$

The equality follows directly from Parseval’s theorem.

Fig 2 shows the nonparametrically estimated graph correlation density $\widehat{\mathcal{C}}(u, v; \mathcal{G})$ for both the bipartite graph ($n = 30$) and political blog ($n = 1490$) examples. The modes represent the higher connectedness. For the bipartite graph, it displays an underlying strong connection between the two disjoint sets and almost no interaction among the nodes from the same group. On the other, hand the web-blog data (directed network) clearly indicates the effect of political polarization. The different sharpness of the two modes indicate the contrasting connection strength within each of the political groups: liberal and conservative. Section 3.2 will further exemplify this point. In addition, note that for small graphs the corresponding shape of graph correlation density is bivariate non-negative (but not necessarily symmetric, as in the case of directed graphs) step-functions while the continuum view (for large graphs as $n \rightarrow \infty$) of $\mathcal{C}(u, v; \mathcal{G})$ is a smooth continuous function over unit interval, hence we can also call it a “correlation field”. Our theory provides a smooth macroscopic description of the microscopic interactions between the nodes (via adjacency matrix) in a compact way.

2.4 Smooth Estimation by Adaptive Filtering

The goal is to select the ‘significant’ LP-transform coefficients $\text{LP}[j, k; \mathcal{G}]$ in a data-driven way. The data driven method of selecting the number of components is backed by substantial literature, see Eubank and Hart (1992), Ledwina (1994) and Inglot et al. (1997). Here, we apply the Ledwina (1994) scheme using Schwarz selection criterion, which is known to be (1) consistent against *every* alternative (including higher-frequency alternatives) to uniformity (Kallenberg and Ledwina, 1995) and (2) enjoys neat asymptotic optimality property (Inglot et al., 2000).

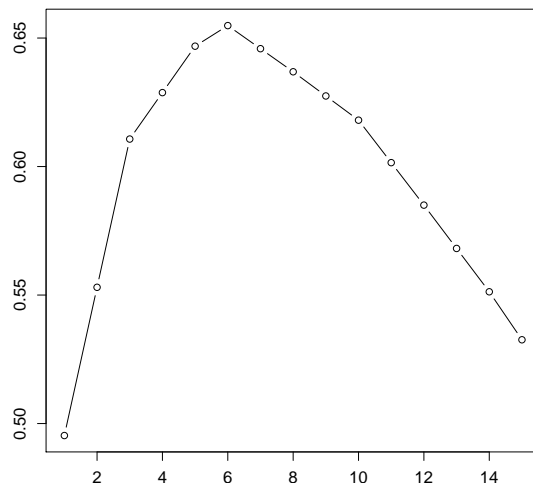


Figure 3: Adaptive filtering in the LP domain for the bipartite graph using the model selection scheme described in Section 2.4.

To identify the important elements using Ledwina’s data-driven model selection criterion, we first rank order the squared LP-coefficients. Then we take the penalized cumulative sum of k coefficients using information criterion $N^{-1} \log(N)k$ and choose the k (number of LP components) for which this is maximum. Fig 3 displays the penalized ordered cumulative sum of LP coefficients for bipartite graph, which selects the number of components $k = 6$.

We estimate the LPINFOR statistic (2.10) by taking the sum of squares of the selected LP-coefficients only. The number of filtered LP transforms (which denote the interesting graph correlations) determine the *effective dimension* of the graph correlation density and tackle the spurious effects to produce smooth estimate $\hat{\mathcal{C}}(u, v; \mathcal{G})$.

3 Applications

3.1 Diagnostic for Testing Null Model

The model building process starts by checking whether the given process comes from the null model. In time series analysis, the flat shape of power spectral density and

the autocorrelation function plot provide a quick and elegant diagnostic tool for the white noise process (which is the null model in the context of time series) that is widely practiced. The following question arises:

Can we propose an analogous portmanteau diagnostic test for null random graph models? What is a reasonable choice of null model for graph data?

In the same spirit as time series analysis, here we propose two (graphical and non-parametric) diagnostic tests based on LP-Fourier transforms and graph correlation density function.

Theorem 3.1 (Asymptotic Normality). *The sample LP graph transforms are i.i.d and*

$$\sqrt{N} \left[\widetilde{\text{LP}}(j, k; \mathcal{G}) - \text{LP}(j, k; \mathcal{G}) \right] \xrightarrow{d} \mathcal{N}(0, 1). \quad (3.1)$$

Define the bivariate score function $L_{jk} := S_j(u)S_k(v) : [0, 1]^2 \rightarrow \mathbb{R}$ and verify $\widetilde{\text{LP}}(j, k; \mathcal{G})$ can be written as $\mathbb{E}[L_{jk}(\tilde{F}(X; X), \tilde{F}(Y; Y)); \tilde{p}]$. The asymptotic normality result then follows immediately by using the same reasoning as given in Ruymgaart et al. (1972).

As a corollary to the previous theorem consider the following interesting fact: The graph LP transforms for random graph \mathcal{G} with probability of an edge equal to $p(x; X)p(y; Y)$ behaves asymptotically normal distribution with mean 0 and variance $1/N$. This result has a threefold significance. First, it tells us the (theoretical) graph correlation density (2.9) is essentially “flat” on the unit square (compare with the power spectral density analogy of time series). Second, it suggests a reasonable definition for null models for graph data, which surprisingly matches the recommendation given in Section 3 of Newman (2006). Third, there is a striking similarity between the asymptotic normality results for sample auto-correlations of white noise process (null model for time series) and our result. One difference is that our higher-order LP graph correlations have the flexibility to detect nonlinear graph connection patterns whereas auto-correlation is limited to linear patterns.

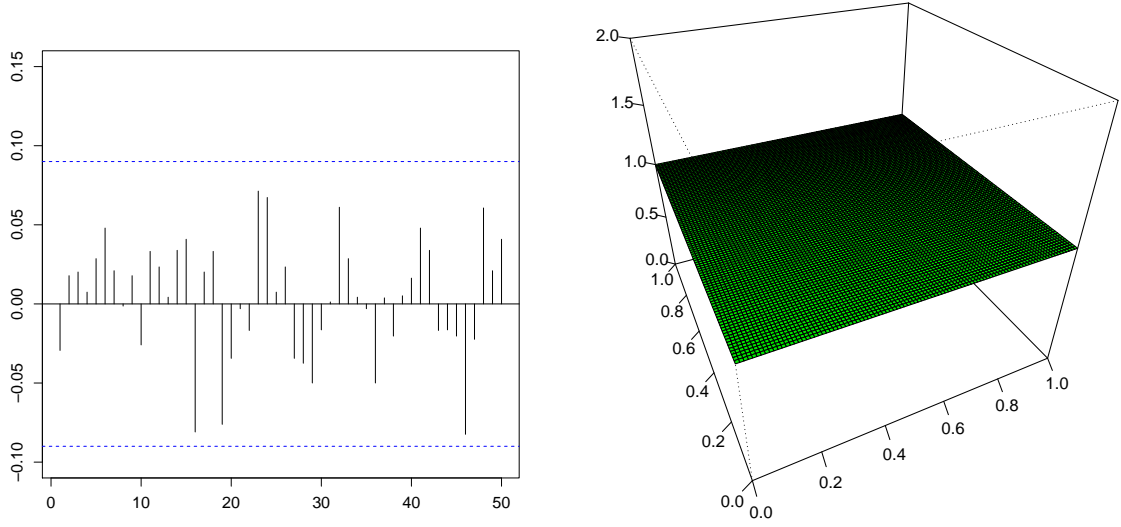


Figure 4: (color online) Graphical diagnostic for testing Erdős-Rényi graph (one specific null model); parallel to white noise testing in time series analysis. (a) LP-graph correlogram; (b) Smooth nonparametrically estimated graph correlation density.

Figure 4 plots the sample LP-transforms or LP-graph correlogram (along with the band $\pm 1.96/\sqrt{N}$, the 95% confidence band) and the *smooth* nonparametrically estimated graph correlation density for the Erdős-Rényi random graph model (Erdős and Rényi, 1959), a member of the null model class. The “flat” or uniform shape of the graph correlation density for Erdős-Rényi model signifies that the *underlying strength of connectedness between all pairs of nodes are the same* (analogous to that in time series the flat spectrum of white noise indicates an equal amount of variance contained in all frequencies). Various other diagnostic statistics for checking random graphs can be developed as a functional of $\mathcal{C}(u, v; \mathcal{G})$; for example, consider the omnibus test statistic $\text{LPINFOR}(\mathcal{G})$, which can be shown to follow chi-square limit distribution under null using Theorem 3.1. It is interesting to note that, our LPINFOR can be viewed as a generalization of Hong (1996) and Box and Pierce’s (1970) type test statistic for graph data.

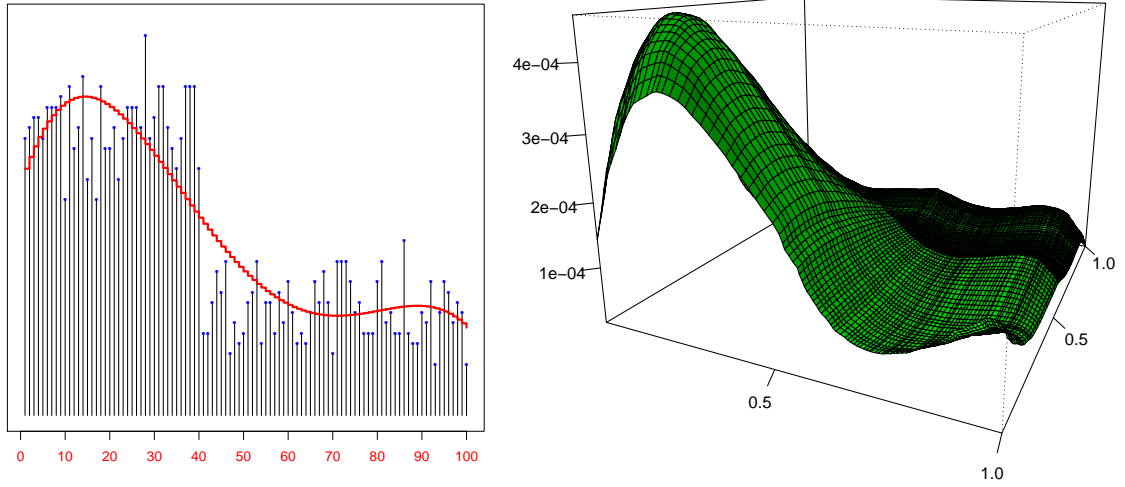


Figure 5: (color online) (a)left: smooth marginal estimate $\hat{p}(x; X)$; (b)right: smooth nonparametric graphon estimate $\widehat{W}(u, v; \mathcal{G})$ of two-block stochastic model (with only one tightly connected group) with sizes 40 and 60 with the edge probability matrix $P = [.6, .1; .1, .1]$

3.2 Smooth Graphon Estimation

Nonparametric modeling of random graphs via Graphon is an emerging frontier of modern random graph theory. We define Graphon (Lovász, 2012) as a bivariate function, $W(u, v; \mathcal{G}) : [0, 1]^2 \rightarrow [0, 1]$ denotes the probability of a directed edge from node x to y , where $Q(u; X) = x$ and $Q(v; Y) = y$. The nonparametric estimation of this theoretical graph object was first initiated by Olhede and Wolfe (2014), and Chan and Airoldi (2014), where they proposed histogram estimates by fitting a stochastic block model (to get the histogram bins by block-averaging). Unfortunately, this neither produces a smooth estimate nor a parsimonious modeling strategy. More specifically, we are interested in the following question:

How can we develop a nonparametric graphon approximation algorithm that will offer smooth estimate via sparse or low-rank modeling?

Our LP estimation scheme makes the first attempt to do so by exploiting the novel

connection between graphon and graph correlation density. The observation that every graphon can be written as a product of null graph model $p(x; X)p(y; Y)$ and the graph correlation density function $\mathcal{C}(F(x; X), F(y; Y); \mathcal{G})$ for $x, y \in V(\mathcal{G})$ permits a new way to estimate graphon, which is applicable for directed/undirected graphs. Besides using nonparametrically estimated $\widehat{\mathcal{C}}(u, v; \mathcal{G})$, one may wish to use smooth estimates of the marginals (instead of raw empirical distributions) $\widehat{p}(x; X)$ and $\widehat{p}(y; Y)$ (expressed as linear combination of discrete LP polynomials as described in the next example).

Consider a two-block stochastic model (with only one tightly connected group) of sizes 40 and 60 with the edge probability matrix $P = [.6, .1; .1, .1]$. Use plug-in smooth nonparametric estimate of the graph correlation density to estimate the graphon. For a more enhanced answer, further smooth the marginals using the model: $p(x; X) = p_0(x; X)[1 + \sum_{j=1}^{n-1} \text{LP}[j; p_0, p]T_j(x; p_0)]$, where $p_0(x; X) = n^{-1}$ is a discrete uniform distribution on the integers $x = \{1, 2, \dots, n\}$. Our specially design orthonormal polynomials $T_j(x; p_0)$ associated with discrete uniform distribution $p_0(x; X)$ exactly match the DLeg_j discrete Legendre polynomials on $(0, 1)$ (which belong to the ‘LP’ family of polynomials). Fig 5(a) shows the fitted (smooth) nonparametric model $\widehat{p}(x; X)$ by selecting the components exactly the same way as discussed in Section 2.4. The resulting LP smooth graphon estimate is shown in Figure 5(b).

Figure 6 shows the estimated asymmetric smooth graphon $\widehat{W}(u, v; \mathcal{G})$ for the political blog data. Bi-modality denotes the presence of two communities. One notable aspect of the shape of the estimated graphon is the contrasting sharpness of the two modes: one is “flat” and the other is more “peaked”; this indicates the heterogeneity of strength of connections within each of the liberal and conservative political web-blogs.

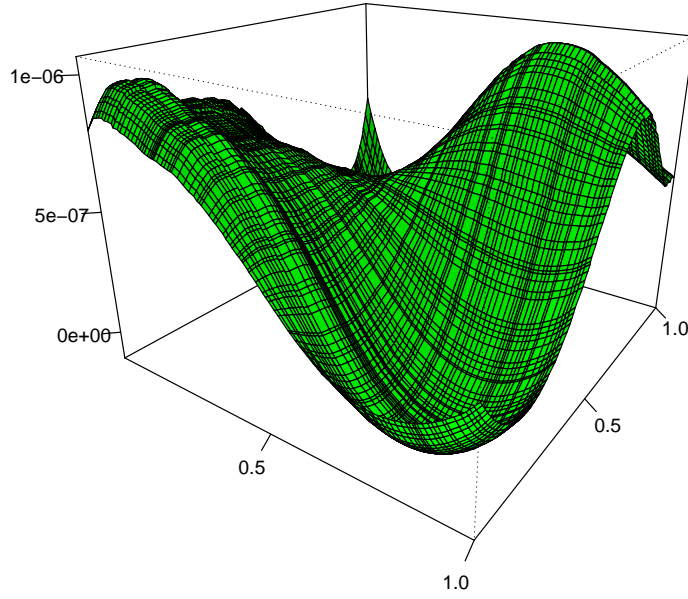


Figure 6: (color online) Smooth nonparametric graphon estimate of the political blog data, which is a directed sparse network.

4 Conclusion

A modern nonparametric modeling approach of graph data is reported here, which has its root in Mukhopadhyay and Parzen (2014). More specifically, this article introduces a new function, called Graph Correlation Density Field, that characterizes the ‘strength of connectedness’ between pairs of vertices of a random graph. Our motivation stems from the desire to develop the concepts of Fourier and inverse-Fourier like transformations for graph data to allow frequency domain interpretation that is similar to time series analysis. To accomplish this, we construct a new class of (data-driven) discrete orthonormal polynomials for graph data. We also discuss the issue of graph compression and sparse representation, which are especially useful for large graphs. We show how these new concepts and tools can be used to solve many outstanding graph modeling problems. In particular, we focus on two problems in this article: (1) developing graphical diagnostic tests for null graph models in the same spirit of testing white noise process in time series analysis. Two diagnostic tools are introduced, which are analogous to the sample auto-correlation function plot and

nonparametric spectral density based white noise testing approach for time series data; (2) the problem of smooth estimation of graphon is discussed by connecting it with graph correlation density. Throughout this paper, we pay special attention to develop ideas and algorithms that are analogous to the time series theory.

We hope this new perspective will enable us to develop a systematic and comprehensive theory of random graphs that is parallel to that of frequency-domain time series theory. The next article of this series will establish some surprising connections between our (modern nonparametric statistics based) LP graph theory and (matrix theory and linear algebra based) spectral graph theory.

Acknowledgment

I would like to express my gratitude to Professor Emanuel Parzen for the encouragement to pursue this research direction. The author also thanks Mohsen Pourahmadi for constructive comments and several useful suggestions.

References

- Adamic, L. A. and N. Glance (2005). The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pp. 36–43. ACM.
- Airoldi, E. M., T. B. Costa, and S. H. Chan (2013). Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. In *Advances in Neural Information Processing Systems*, pp. 692–700.
- Box, G. and D. Pierce (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association* 65(332), 1509–1526.
- Brockwell, P. J. and R. A. Davis (2009). *Time series: theory and methods*. Springer.

- Chan, H. S. and E. M. Airoldi (2014). A consistent histogram estimator for exchangeable graph models. *Journal of Machine Learning Research Workshop and Conference Proceedings* 32(1), 208–216.
- Chatterjee, S. (2014). Matrix estimation by universal singular value thresholding. *The Annals of Statistics* 43(1), 177–214.
- Erdős, P. and A. Rényi (1959). On random graphs. *Publ. Math. Debrecen* 6, 290–297.
- Eubank, R. and J. D. Hart (1992). Testing goodness-of-fit in regression via order selection criteria. *The Annals of Statistics*, 1412–1425.
- Frieze, A. and R. Kannan (1999). Quick approximation to matrices and applications. *Combinatorica* 19(2), 175–220.
- Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica: Journal of the Econometric Society*, 837–864.
- Inglot, T., W. C. M. Kallenberg, and T. Ledwina (1997). Data driven smooth tests for composite hypotheses. *Annals of Statistics* 25, 1222–1250.
- Inglot, T., W. C. M. Kallenberg, and T. Ledwina (2000). Vanishing shortcoming and asymptotic relative efficiency. *The Annals of Statistics* 28(1), 215–238.
- Kallenberg, W. C. and T. Ledwina (1995). Consistency and monte carlo simulation of a data driven version of smooth goodness-of-fit tests. *The Annals of Statistics*, 1594–1608.
- Ledwina, T. (1994). Data driven version of neyman smooth test of fit. *Journal of the American Statistical Association* 89, 1000–1005.
- Lovász, L. (2012). *Large networks and graph limits*, Volume 60. American Mathematical Soc.
- Lovász, L. and B. Szegedy (2006). Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B* 96(6), 933–957.

- Mukhopadhyay, S. and E. Parzen (2014). LP approach to statistical modeling. *Unpublished Technical Report available at arXiv:1405.2601*.
- Newman, M. E. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical review E* 74(3), 036104.
- Olhede, S. C. and P. J. Wolfe (2014). Network histograms and universality of block-model approximation. *Proceedings of the National Academy of Sciences* 111(41), 14722–14727.
- Ruymgaart, F., G. Shorack, and W. van Zwet (1972). Asymptotic normality of nonparametric tests for independence. *The Annals of Mathematical Statistics*, 1122–1135.